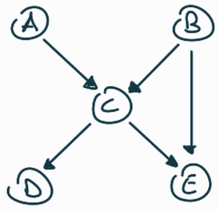
Bayesian Networks

We can now solve problems in which we have
- Dataset $D = \{(x_n)\}_{n=1}^{N}$
- Hidden variables Z

We assumed there is a dependence between X and Z and X can be expressed from Z with $P(X|Z)$.

We can define a relation between X and Z. This is a Bayesian Network with just two random variables, X and Z, in which we make explicit that X depends on Z.

This can be generalized to cases in which we have several random variables. We know from domain knowledge that some variables are affected by other variables.



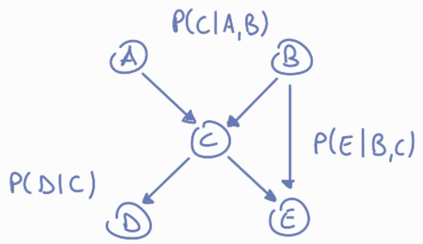
A probabilistic graphical model (PGM) is a graph representing conditional probability distributions among random variables. If we consider discrete random variables we can express each of this edges in form of a conditional probability table.

$B, E \in \{0,1\}$ (boolean variables)

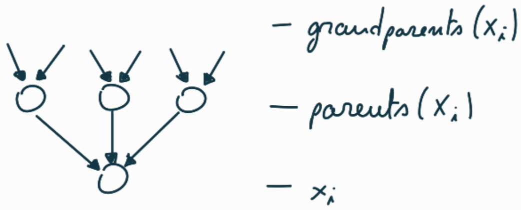
$$B \xrightarrow{P(E|B)} E$$

	$B=0$	$B=1$
$E=0$	0.8	0.7
$E=1$	0.2	0.3
	$\Sigma=1$	$\Sigma=1$

probability of E being 0 given that B is also 0



x_i is conditional independent from all the variables that are not its direct parents



$$P(x_i \mid parents(x_i))$$

Conditional independence means that $P(x_i \mid parents(x_i), Z) = P(x_i \mid parents(x_i))$

without Bayesian Network

$P(A,B,C,D,E)$: we had to compute the joint probability

with Bayesian Network

$P(A)$ → random variables with no parents
$P(B)$

$P(C|A,B)$
$P(E|C,B)$
$P(D|C)$

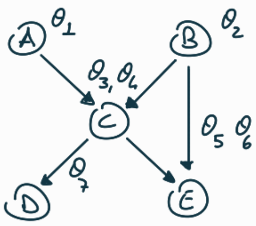
} each with its conditional probability table

Once we have these parameters we can compute any information from this model, we can do inference

We can apply the chain rule to $P(A,B,C,D,E)$

$$P(A,B,C,D,E) = P(A|B,C,D,E)\,P(B|C,D,E)\,P(C|D,E)\,P(D|E)\,P(E)$$

We can consider the values of each conditional probability table as a parameter and the whole Bayesian Network as the model for our problem.



parametric model → $\theta = \langle \theta_1, \dots, \theta_n \rangle$

It's easy to solve this kind of problem if all the variables are observable and present in the dataset. What if some variables are completely non-observable?

Very simple if $D = \{ (a_n, b_n, c_n, d_n, e_n) \}_{n=1}^{N}$
What if $D = \{ (c_n, d_n, e_n) \}_{n=1}^{N}$ not all the variables observable

This can be modeled with hidden variables

Remarks

EM algorithm is good when the input space has low dimension.

$X \in \mathbb{R}^{d}$ → d should be low

If we are dealing with images, $X \in \mathbb{R}^{w \times h \times d}$ w = width, h = height, d = depth / channels of the image
EM is not suitable.